



Fusion of stereo vision and Time-Of-Flight imaging for improved 3D estimation

Guðmundsson, Sigurjón Árni; Aanæs, Henrik; Larsen, Rasmus

Published in:

International Journal on Intelligent Systems Technologies and Applications (IJISTA)

Link to article, DOI:

[10.1504/IJISTA.2008.021305](https://doi.org/10.1504/IJISTA.2008.021305)

Publication date:

2008

Document Version

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):

Guðmundsson, S. Á., Aanæs, H., & Larsen, R. (2008). Fusion of stereo vision and Time-Of-Flight imaging for improved 3D estimation. *International Journal on Intelligent Systems Technologies and Applications (IJISTA)*, 5(3/4), 425 - 433. <https://doi.org/10.1504/IJISTA.2008.021305>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Fusion of stereo vision and Time-Of-Flight imaging for improved 3D estimation

Sigurjón Árni Guðmundsson*, Henrik Aanæs
and Rasmus Larsen

Technical University of Denmark,
Informatics and Mathematical Modelling,
Richard Petersens Plads, Building 321,
2800 Kgs.Lyngby, Denmark
E-mail: sag@imm.dtu.dk
E-mail: haa@imm.dtu.dk
E-mail: rl@imm.dtu.dk
*Corresponding author

Abstract: This paper suggests an approach in fusing two 3D estimation techniques: stereo vision and Time-Of-Flight (TOF) imaging. By mapping the TOF-depth measurements to stereo disparities, the correspondence between the images from a fast TOF-camera and standard high resolution camera pair are found, so the TOF depth measurements can be linked to the image pairs. In the same framework, a method is developed to initialise and constrain a hierarchical stereo matching algorithm. It is shown that in this way, higher spatial resolution is obtained than by only using the TOF camera and higher quality dense stereo disparity maps are the results of this data fusion.

Keywords: data fusion; real-time 3D computer vision; stereo vision; Time-Of-Flight; TOF.

Reference to this paper should be made as follows: Guðmundsson, S.Á., Aanæs, H. and Larsen, R. (xxxx) 'Fusion of stereo vision and Time-Of-Flight imaging for improved 3D estimation', *Int. J. Intelligent Systems Technologies and Applications*, Vol. x, No. x, pp.xx-xx.

Biographical notes: Sigurjón Árni Guðmundsson received a BSc in Electrical and Computer Engineering from the University of Iceland (UI) in 2001 and an MSc from the Informatics and Mathematical Modelling at the Technical University of Denmark (DTU) in 2006. Currently, he is pursuing a PhD from UI and working as a Research Assistant at DTU. His areas of interest lie within the field of computer vision and machine learning.

Henrik Aanæs received his PhD from the Informatics and Mathematical Modelling at the Technical University of Denmark in 2003, where he currently holds the position of Assistant Professor (Lektor) in Computer Vision. His current research efforts are on statistical inference from imaging devices, in particular, the estimation of 3D shape.

Rasmus Larsen received his MSc (Eng) and PhD from the Informatics and Mathematical Modelling at the Technical University of Denmark. He worked as an Expert Statistician with Novo Nordisk A/S and currently holds a position as an Associate Professor in Medical Image Analysis at the Technical University of Denmark. His research interests include biomedical image analysis and multivariate statistics.

1 Introduction

3D computer vision is a very active research field. Research driven by the need for solutions in such broad fields as robot vision, automatic navigation, computer interaction, tracking and action recognition, only to name a few. Here, a method is described that obtains quality 3D estimates using a fusion of the traditional and thoroughly investigated depth from stereo image technique and the new technology of Time-Of-Flight (TOF) range cameras. The three main contributions of this paper are: a disparity estimate derived from the TOF-depth measurements is found that accurately describes the correspondences between the TOF-and standard camera. Secondly, a method to transfer the TOF-depth to a standard image via the disparity estimate. Finally, a disparity estimate is used to initialise and constrain a stereo matching algorithm.

The results show that higher spatial land depth resolution is gained using this fusion than by only using the TOF camera and more correct and higher quality dense stereo disparity maps in difficult scenarios than using standard stereo methods.

In the remainder of this section, the theoretical background is given and some of the related work presented. In Section 2, the fusion methods are presented. Section 3 introduces the set-up with the TOF-stereo rig and shows the results of the fusion methods and the stereo matching algorithm. Finally, some conclusions are drawn in Section 4.

1.1 Depth from stereo

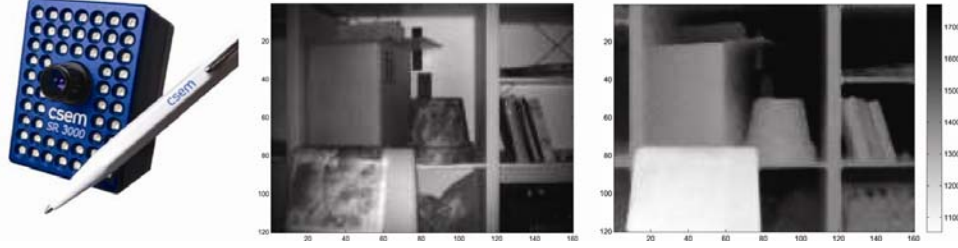
3D from stereo strives at computing corresponding pixels for every pixel of an image pair. The correspondence can be expressed as a disparity vector, i.e. if the corresponding pixels are x_l and x_r in the left and right image, respectively, then the disparity map $D_l(x_l, x_r)$ is the difference of their image coordinates. Therefore, the output of a stereo algorithm is therefore this disparity map that maps every pixel from one image to the other. The standard approach is to simplify the disparity vector to one dimension. This is done by performing rectification where the image pairs are transformed by 3×3 homographies, which align epipolar lines to the corresponding scanlines. A detailed description of the two view epipolar geometry can be found in Hartley and Zisserman's book (2004).

Most stereo algorithms work by minimising a cost function and results of various research on such algorithms are reviewed in Dhond and Aggarwal (1989) and are further compared in Scharstein and Szeliski (2002). The stereo algorithm used here is a very effective, hierarchical, Dynamic Programming (DP) approach where the matches are made in sequences under order constraints and solved as a global optimisation problem. The algorithm used here is based on Van Meerbergen et al.'s work (2002).

1.2 Depth from Time-Of-Flight

The TOF camera used here is a SwissRanger SR3000.¹ It is designed to be a cost-efficient and eye-safe range image solution.

Basically, the SR3000 is an amplitude modulated light source and a 2D sensor built in a miniaturised package (see Figure 1).

Figure 1 The SwissRanger SR3000 and its output: the intensity image I_t and the range image Z_t 

The sensor is a 176×144 pixel custom designed $0.6 \mu\text{m}$ CMOS/CCD chip where each pixel in the sensor demodulates the reflected light by a lock-in pixel method (Lehmann et al., 2005). The sensor reconstructs the returning signal and two images are generated: an intensity (grey scale) image derived from the amplitude of the signal and a range image (depth measurement per pixel) derived from the phase offset of the signal. Figure 1 shows images from the camera.

The SR3000 has its pros and cons. The camera's low latency makes it very attractive for many applications as it can generate around 20 frames a second and has the depth accuracy of approx. 1 cm at 1 m which is sufficient for various applications (Oggier, Büttgen and Lustenberger, 2005). On the other hand, the limited spatial resolution of the images and low quality intensity image make many computer vision techniques difficult to practise. This is why it is important to develop methods to fuse the information from the TOF camera with other data; e.g. standard high resolution cameras and multiple view depth measurement set-ups.

1.3 Data fusion

The TOF-sensor technology is a very new technology and thus not much has been done in the field of data fusion. PMD technologies² is another producer of TOF-sensors with which some experiments in data fusion have been done recently. Beder, Bartczak and Koch (2007) have done some benchmark comparisons between TOF-measurements and stereo depth measurements using a rig with both a TOF camera and standard cameras. Also, Kuhnert and Stommel (2006) experimented with a set-up of a stereo rig and TOF camera. Here, they took the two measurements and combined them knowing their relative positions and choosing the measurement that gives better results regionally.

2 Methods

Two fusion methods are proposed here both in a calibrated set-up. First, the TOF depth measurement Z_t is fused with one standard image I_s by estimating the disparity and assigning the Z_t points to the corresponding points in I_s . Secondly, by using Z_t to constrain a stereo algorithm per pixel. This is also done by estimating a disparity between the corresponding points between Z_t and both I_l and I_r .

2.1 Fusing Z_t with I_s

Depth and disparity are interconnected in standard parallel stereo by:

$$Z = \frac{Tf}{D} \quad (1)$$

where Z is the depth in Euclidian coordinates, T is the baseline, f is the focal length and D is the disparity.

Looking at the rectified I_s - Z_t reference frame, Z_t can be mapped to the disparity $D_t(t, s)$, where the TOF-camera is the reference using the relevant baseline and focal lengths getting:

$$D_t(t, s) = \frac{T_{st}f}{H_{st}^t Z} \quad (2)$$

where H_{st}^t is the homography to rectify the TOF image in the standard image-TOF pairing. This disparity estimate can now be used to assign the TOF-depth directly to the standard image. This disparity estimate is in the TOF-camera's low resolution so for a high resolution solution interpolation is needed. Results of this merging of TOF and standard images are presented in Section 3.2.

2.2 Disparity estimate for TOF-stereo

Using Z_t in a set-up with two standard images I_l and I_r is more complicated. The method we employ is based on a method named correspondence linking (Koch, Pollefeys and Van Gool, 1998). The corresponding left and right points, x_l and x_r , have to be linked through the TOF-image which is in between I_l and I_r . This means that for each pixel x_t in the Z_t the corresponding pixels in I_l and I_r found by:

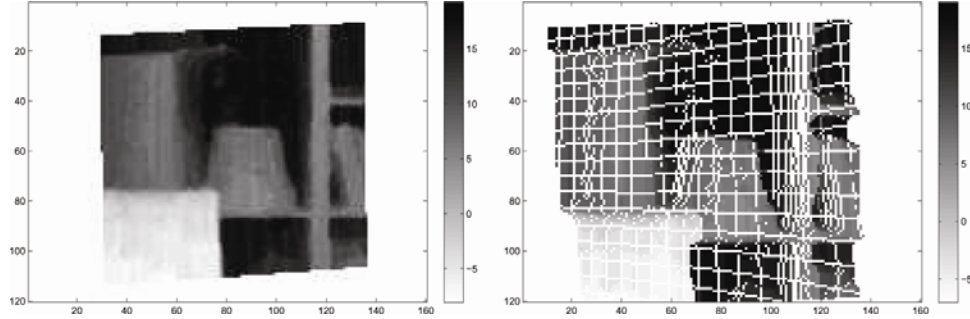
$$\begin{aligned} x_r &= (H_{tr}^r)^{-1} D_t(t, r) H_{tr}^t x_t \\ x_l &= (H_{tl}^l)^{-1} D_t(t, l) H_{tl}^t x_t \end{aligned}$$

where H_{tr}^r and H_{tr}^t are the homographies to rectify the right and TOF-image, respectively, in the I_r - I_t reference frame and $D_t(t, r)$ is the function that maps the disparity in this frame. The second line does the same in the I_l - I_t frame.

Calculating x_r and x_l for each point x_t in the TOF-image and transforming them to the left-right rectified frame gives the 1D disparity between the left and right images:

$$D_t(l, r) = H_{lr}^l x_l - H_{lr}^r x_r \quad (3)$$

$D_t(l, r)$ is now defined in the TOF-image's reference frame and needs to be resorted by the left image coordinates resulting in a disparity map estimate for the left reference frame $D_l(l, r)$. Figure 2 shows the results and how the $D_t(l, r)$ 'blows up' when resorted, i.e. there are points in the standard images that have no corresponding points in the TOF image. This is caused by the difference in focal length of the two cameras; they have been scaled to 160×120 , but the lenses are very different. The grid of no values is filled up by interpolation.

Figure 2 The Time-Of-Flight disparity estimate $D_l(l, r)$ 

Notes: First in the TOF-camera's reference frame and then rearranged for the left camera. As the focal length of the standard camera is larger than the TOF-camera's the data 'blows-up'. These holes were filled by interpolation as seen in the top left image in Figure 5.

2.3 Constrained stereo matching

The TOF-disparity estimate $D_l(l, r)$ is now used in a hierarchical stereo matching DP algorithm mentioned in Section 1.1. $D_l(l, r)$ is placed as the initial fourth level of the hierarchy, and in that way is used as a 'offset' disparity map to the next higher level whose results are used in the same way in the next level and so on. This way the algorithm has an initial per pixel constraint on the disparity search space instead of having just an overall minimum–maximum disparity range initialisation as in the standard case. Results are shown in Section 3.3.

3 Set-up and results

3.1 Design and calibration of the TOF-Stereo rig

The TOF-stereo rig shown in Figure 3 is designed to acquire depth resolution of around 1 mm at 1 m and have a sufficient overlap of the left and right images to be used at 90 cm – 4 m. For this, the baseline was chosen 30 cm and the cameras verge towards each other at 7° each from the parallel set-up.

Figure 3 The Time-Of-Flight stereo rig and the rectified image pair I_l and I_r used in the experiment

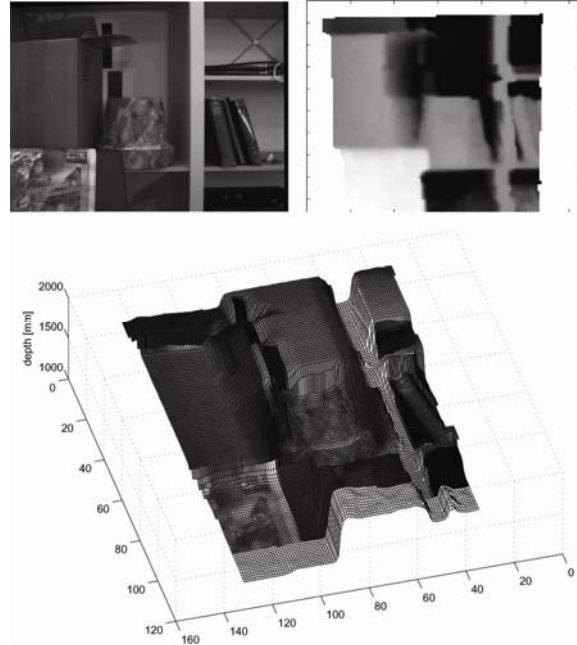
Calibration quality is essential for successful stereo vision and here even more so when there are three cameras with one having such low spatial resolution. All the three cameras were calibrated and stereo-calibrated using Bouguet's Matlab calibration toolbox.³ The TOF-images were cropped to 160×120 to have the same format as the standard cameras.

The TOF camera was stereo-calibrated with each eight times down-scaled left and right image. The left and right were then calibrated with each other in full size (1280×960). The six rectification homographies needed in the algorithm (in Section 2) were found by using the method described in Fusiello, Trucco and Verri (2000).

3.2 Fusing Z_t with I_s

Figure 4 shows the results of linking Z_t to a single high resolution image I_s (here I_l) via $D_r(t, s)$. The results have artifacts directly inherited from the TOF-camera, i.e. the quite low depth resolution and noise. For example, the edge on the flower pot can barely be detected. Also, the rounding of the estimated disparity points give some erratic behaviour close to depth discontinuities but the over all results are not bad.

Figure 4 The results of the Time-Of-Flight left image fusion



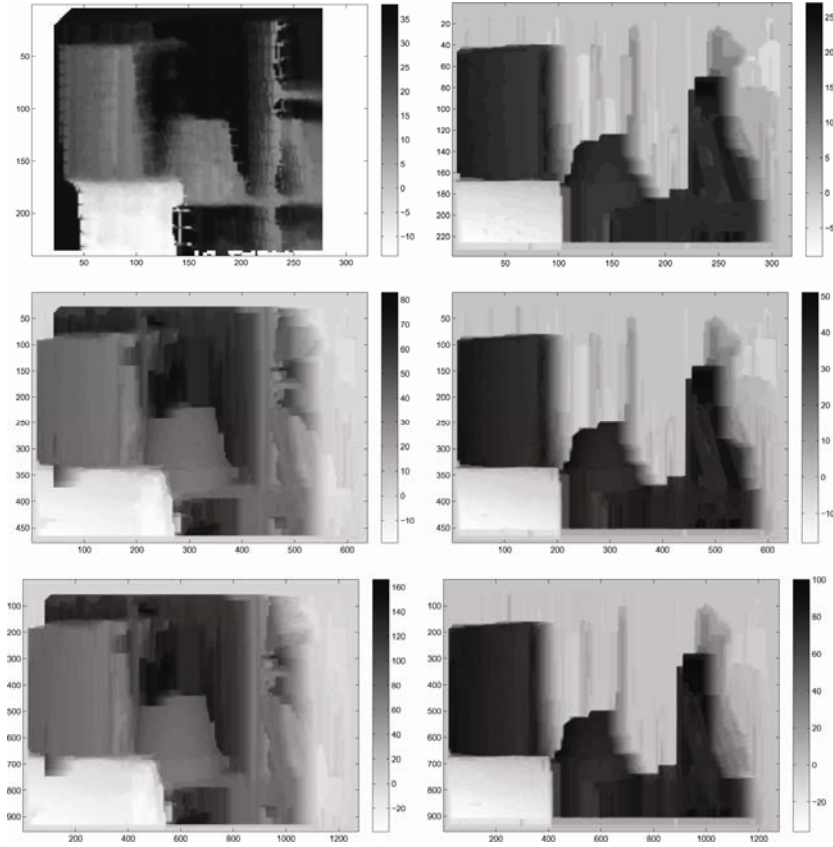
Notes: Top: the left image and its depth map. Bottom: geometric reconstruction in low resolution shown from a higher angle.

When the depth map is up-scaled and assigned to the full-size image, there are more problems at edges that are overly smooth the results still have the artifacts typical to the TOF-depth measurements. This method still gives a good fast but somewhat crude depth estimate on high resolution images which can suit well for many applications.

3.3 Constrained stereo matching

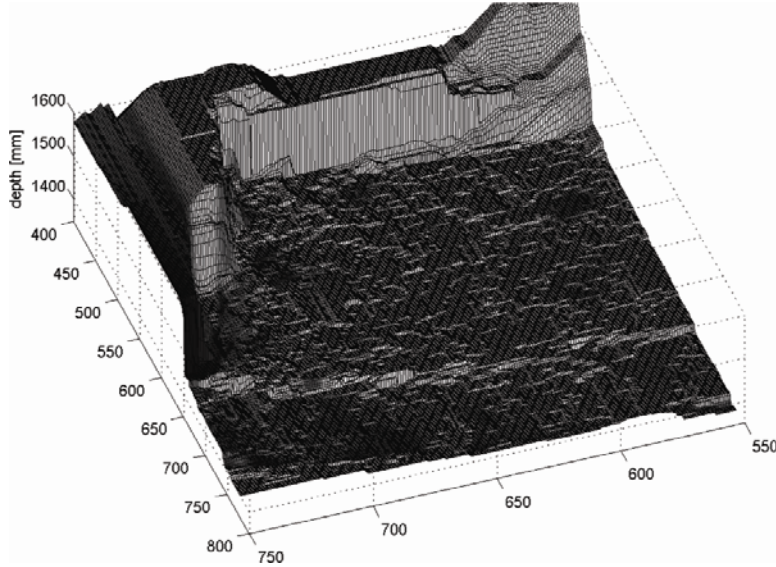
Figure 5 illustrates the results of the disparity calculations on different levels of the hierarchy; the left column when using $D_l(l, r)$ as input on the fourth level and for comparison the right column shows results from the standard method when the algorithm is initialised with a single disparity range for the whole image, here simply given by the minimum–maximum range of $D_l(l, r)$. The results using $D_l(l, r)$ per pixel are much better giving correct disparities at difficult areas like the uniform back wall where the standard approach fails totally. Also, the results are smoother and more detailed than the standard case. No improvements are made in the case of the standard method if it is run with a higher level pyramid. Figure 6 shows a cut out of the reconstructed geometry of the flower pot. The edge is very clear and it gives a smoother and more correct result than from the simple fusion results in Section 3.2.

Figure 5 Results from different levels of the hierarchical Dynamic Programming matching



Notes: Left column: initialised with $D_l(l, r)$ (top), then two higher levels of the pyramid. Right column: standard DP method initialised with the range of $D_l(l, r)$. The results from three increasing hierarchy levels. The final results on the bottom left are much superior to the one on the bottom right. For example, the standard DP algorithm totally fails at matching correctly the uniform areas like the wall in the back (which appears to be in front of the flower pot).

Figure 6 The reconstructed geometry of the flower pot from the final disparity map using the $D_I(l, r)$ estimate



4 Conclusions

An algorithm has been presented to fuse TOF images with images from standard camera pairs in a calibrated framework. The results are promising especially with the TOF-constrained stereo matching algorithm showing much higher quality disparity maps, in a difficult scenario, than from the standard approach. Here, the TOF information has only been used as an initialisation or bootstrapping of the DP algorithm but one can imagine that a fuller fusion can be achieved. One way would be, e.g. by backwards warping with lookup interpolation from the stereo-algorithms results to manage better occlusions and to avoid artifacts such as those depicted in Figure 2. Also, by adding the disparity estimate into the DP cost-function as a further conditioner would possibly make this fusion stronger and exploiting better the available data for improved 3D estimation.

Acknowledgements

This work was in part financed by the Graduate School for Biomedical Optics and New Laser System under a grant from the Danish Ministry of Science, Technology and Innovation. Also by the EU framework 6 IST project ARTTS.

References

- Beder, C., Bartczak, B. and Koch, R. (2007) 'A comparison of pmd-cameras and stereo-vision for the task of surface reconstruction using patchlets', *The Second International ISPRS Workshop, Ben COS*.
- Dhond, U.R. and Aggarwal, J.K. (1989) 'Structure from stereo-a review', *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, pp.1489–1510.
- Fusiello, A., Trucco, E. and Verri, A. (2000) 'A compact algorithm for rectification of stereo pairs', *Machine Vision and Applications*, Vol. 12, pp.16–22.
- Hartley, R.I. and Zisserman, A. (2004) *Multiple View Geometry in Computer Vision* (2nd ed.). Cambridge, UK: Cambridge University Press, ISBN: 0521 54051 8.
- Koch, R., Pollefeys, M. and Van Gool, L. (1998) 'Multi viewpoint stereo from uncalibrated video sequences', Paper presented in the *Proceedings ECCV'98, LNCS* (Vol. 1406, pp.55–71).
- Kuhnert, K.D. and Stommel, M. (2006) 'Fusion of stereo-camera and pmd-camera data for real-time suited precise 3D environment reconstruction', *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference*, pp.4780–4785.
- Lehmann, M., Kaufmann, R., Lustenberger, F., Büttgen, B. and Oggier, T. (2005) 'CCD/CMOS lock-in pixel for range imaging: challenges, limitations and state-of-the-art', H. Ingsand and T. Kahlmann (Eds), Paper presented in the *Proceedings of the First Range Imaging Research Day* (pp.21–32), Zurich.
- Oggier, T., Büttgen, B. and Lustenberger, F. (2005) 'Swissranger sr3000 and first experiences based on miniaturized 3d-TOF cameras', H. Ingsand and T. Kahlmann (Eds), Paper presented in the *Proceedings of the First Range Imaging Research Day* (pp.97–108), Zurich.
- Scharstein, D. and Szeliski, R. (2002) 'A taxonomy and evaluation of dense two-frame stereo correspondence algorithms', *Int. J. Computer Vision*, Vol. 47, pp.7–42.
- Van Meerbergen, G., Vergauwen, M., Pollefeys, M. and Van Gool, L. (2002) 'A hierarchical symmetric stereo algorithm using dynamic programming', *Int. J. Computer Vision*, Vol. 47, pp.275–285.

Notes

¹<http://www.swissranger.ch>.

²<http://www.pmdtec.de>.

³http://www.vision.caltech.edu/bouguetj/calib_doc/.